What Makes a Good Paper, and What Makes a Good Talk? Two Seminar Sessions and Their Checklists*

Roman Klinger

2026-11-13

In the winter term 2026/2027, I teach the master's seminar *Large Language Models for Natural Language Understanding* at the University of Bamberg. In this seminar, each student selects a research paper on a topic of their choice and presents it to the group, in a 25-minute talk followed by 10 minutes of discussion. Before the first student talk, we spent two sessions on two questions that matter for this setup: How do you decide whether a paper is worth presenting? And what makes a good talk?

In both sessions, I did not hand out a list of rules. Instead, the students worked out the criteria themselves in small groups, and at the end of each session we compiled a joint checklist. After the sessions, I edited and structured these lists. I think they are useful beyond the seminar, so I share them here, together with the material for both sessions, in case you want to run a similar exercise.

Session 1: What makes a good paper?

The students worked in groups of up to four people. Each group assigned two roles: one person kept track of time, and one person made sure that the group ended up with a list of statements and presented it later. Every group received a research paper from the topic area of the seminar, as PDF and as a printout, and assessed it in two phases, guided by a sheet with questions.

- **Phase 1 (10 minutes): a quick assessment.** Read only the title, the abstract and the conclusion, and look at the author list, the figures, the layout and the publication venue. Does the paper deserve more attention? Collect five arguments for and five arguments against reading it further.
- **Phase 2 (20 minutes): a closer reading.** Read the paper in more detail, not every line, but enough to understand what happens. The sheet guides this in four steps: the problem formulation and novelty (What is the research question? Why does it matter? What is the contribution, and is it new?), the methodological soundness (Does the related work make the starting point clear? Could others replicate the method?), the results and conclusions (Do the evaluations support the claims?),

*Published as blog post at https://www.romanklinger.de/blog/2026-11-13-paper-and-talk-quality/

and finally an overall decision, as if the group were reviewing the paper: accept or reject?

While working on the paper, everybody noted statements of the form “A good paper does …”. In the remaining 30 minutes, the groups first consolidated their notes, and then we collected all statements into one checklist.

The immediate goal was practical: the students had to choose the paper that they would present, and the checklist helps with that decision. But the exercise also trains how to give feedback on a paper, and, later, how to write a good paper yourself.

Material: assessment sheet

Checklist: What makes a good paper?

Use it to decide whether to read a paper, to review a paper, or to check your own writing. Not every item applies to every type of paper (empirical study, theoretical work, survey, position paper, user study).

Signals beyond the content These signals are neither sufficient nor always reliable. They can, however, help to narrow down an otherwise too large set of papers before you decide what to read.

- The paper is published at a good venue (conference or journal).
- The paper has been peer reviewed; the type of review is known (e.g., double-blind, or open review with public reviewer comments).
- The venue has a good reputation for its review process. Note that a prestigious name does not guarantee rigor.
- Some of the authors are at renowned institutions.
- The author list includes established researchers, ideally with a track record in the specific subfield.
- Preprints (for instance on arXiv) have not been peer reviewed and should be read with caution. Check what changed between versions.
- The paper has been cited frequently (taking into account how long it has been available).
- Others have replicated, reproduced, or critiqued the work.
- Funding sources and conflicts of interest are disclosed.
- There are no retractions or unaddressed errata.

Presentation

- The title is clear.
- The abstract reflects the content accurately and does not overclaim.
- The figures are informative.
- Tables and figures can be understood on their own (captions, axis labels, units, legends).
- Uncertainty is visualized (error bars, confidence intervals), not only single numbers.
- An overview figure makes the method easy to grasp.
- The data that is used is presented clearly.

- Terminology and notation are consistent throughout the paper.
- Every claim in the text can be traced back to specific evidence (a figure, a table, a section).
- The results are understandable; the text guides readers through their interpretation.
- Important words and numbers are highlighted. (This helps skimming, but it can also mislead if it emphasizes cherry-picked results.)

Motivation and problem definition

- The research question or hypothesis is formulated clearly.
- The problem is well motivated: Why does it matter, and for whom?
- The scope of the problem is appropriate, neither trivial nor unfocused.
- Assumptions are stated explicitly.
- The contribution is stated clearly, ideally as a short list, and can be checked against the content of the paper.
- Research question, contribution, and conclusion are clearly connected.

Related work

- The related work is comprehensive and well structured.
- The paper states explicitly how it differs from previous work.
- Recent work is included, not only the authors' own previous papers.
- Foundational work is acknowledged.
- Competing approaches are treated fairly, including those that perform better than the proposed method.

Research methods *Design and justification*

- The reasons for the methodological choices are clear.
- The methods are not more sophisticated than necessary: a simple method that works is better than a complicated one with similar performance.
- The experimental setting allows drawing the conclusions that the paper draws.
- There are no obvious mistakes in the methodology, such as evaluating on the training data or using a statistical test whose assumptions do not hold.
- The paper guards against data leakage (information from the test data influencing training or model selection), repeated use of the same test set during development, and cherry-picking of results or datasets.

Data

- Multiple datasets are used for the evaluation.
- The datasets are representative of the real-world setting, and their limitations are discussed (size, bias, collection process, label quality).

Evaluation

- The baselines are strong and appropriate, including simple ones and the current state of the art.

- The baselines are tuned with comparable effort (hyperparameter search, compute) as the proposed method.
- The evaluation metrics are appropriate and justified, ideally more than one.
- Ablation studies show which components of the method actually contribute.
- The statistics are sound: multiple runs with different random seeds, reported variance, significance tests where appropriate, and corrections when many comparisons are made.
- Qualitative analyses and error analyses complement the aggregated numbers.
- Robustness and generalization are examined (other domains or data distributions, sensitivity to hyperparameters, behavior with model size).
- Computational cost, runtime, and resource requirements are reported.

Reproducibility

- Code and data are available to reproduce the study.
- The code is documented, runnable, and versioned.
- The description of the experimental setting (possibly in the appendix) is detailed enough to reproduce the work, including random seeds, software versions, and hardware.

Studies with human participants and qualitative research

- User studies justify the number of participants and describe how they were recruited; they report ethics approval and, where relevant, pre-registration (publishing the study plan before collecting data).
- Qualitative research describes how the material was coded, reports the agreement between annotators, and reflects on the researchers' own influence on the results.

Results and interpretation

- The results are reported honestly, including negative or unexpected findings.
- Improvements are meaningful in practice, not only statistically significant or marginal.
- The claims are proportionate to the evidence (e.g., no causal language for correlations).
- Alternative explanations for the results are considered.

Conclusion, discussion, and limitations

- The conclusion is concrete and concise.
- The conclusions match the motivation.
- The paper is transparent about what the study cannot show; limitations are made explicit.
- Failure cases and negative results are shown.
- Broader implications and practical applicability are discussed.
- Future work is concrete rather than generic.
- Societal impact, ethical considerations, and potential misuse are addressed where relevant.

Overall quality

- Novelty or significance: the paper advances understanding, offers a new perspective, or provides a useful negative result.
- Internal consistency: text, figures, and tables do not contradict each other.
- Writing quality: the paper is well structured, readable, and free of unnecessary jargon.
- Appropriate length: neither padded nor too compressed to follow.

Session 2: What makes a good talk?

This session started with two short talks that I gave myself. Both presented the same (fictional) study, a comparison of prompted large language models with fine-tuned models for emotion classification. All numbers in both talks are invented.

The first talk was deliberately bad. It started with a generic outline and a wall of text, never stated a research question, used “emotion” and “sentiment” interchangeably, hid the main finding at the bottom of a large results table (where it contradicted the bar chart on the same topic), and rushed through the results at the end. I also mostly faced the screen, spoke too fast, let my phone ring, had notifications popping up, and included a picture of a cat that had nothing to do with the content. The second talk presented the same study the way I would want it: it started with an example, stated the question, put the main finding into the title of the results slide, and ended with take-home messages. It still contained two small deliberate flaws (an unexplained evaluation measure and a slide with more text than necessary), to see whether the students would notice them.

During both talks, the students took notes on an observation sheet, organized along aspects such as the opening, the storyline, the level of detail, slide design, figures, delivery, time management, and the ending. Afterwards, they discussed their notes in groups, as in the previous session, and wrote down statements of the form “A good talk …”. Finally, we compiled a joint checklist.

Material: observation sheet, the bad example talk with its speaker guideline (what to do wrong on which slide, and what to reveal afterwards), the good example talk with its speaker guideline

Checklist: What makes a good talk?

Use it to prepare and rehearse your own talk, and to give feedback to others. It was written for a 25-minute talk in which you present a research paper, followed by 10 minutes of discussion, but most points apply to scientific talks in general. Not every item applies to every talk.

Preparation

- Set up the presentation beforehand: test the projector, the adapter, and the presenter view; have a PDF backup on a USB stick.
- Turn off distractions: phone, notifications, other open windows.
- Rehearse the talk out loud at least once, with a timer.

- Understand the content that you present; be able to explain every slide without reading it.
- Plan your time: roughly one slide per minute. Decide in advance what to cut, not during the talk.
- Make the slides yourself instead of letting a language model generate them. Generated slides tend to add pathos, use overly metaphoric language, contain inaccuracies, and oversimplify, and without experience in giving talks, you may not notice these issues. Using AI for clearly specified parts is fine, for instance a figure that you describe in detail, but write the texts on the slides yourself and decide yourself which content ends up in the talk. (I explain this in more detail in my thesis checklist post.)

Opening

- The title communicates the main content of the talk.
- The title slide contains the author information (when you present someone else's paper: the paper's authors and venue, and your own name).
- The talk starts with a catching, clear motivation: Why should the audience care?
- It starts with a concrete example or question rather than with a generic outline. A standard "Introduction, Method, Results, Conclusion" outline is typically not helpful.
- The research question is stated explicitly and early.

Structure and storyline

- There is a clear thread: each slide follows from the previous one, and the audience always knows where they are.
- The talk does not jump back to topics from earlier; order the content so that this is not necessary.
- Each slide has one main message, and the slide title fits the content (ideally, it states the message).
- The related work describes the starting point of the presented work and helps to understand the talk.
- When you present someone else's paper: separate clearly what the authors did and claim from your own assessment.

Content and level of detail

- The content fits the audience and is understandable for them.
- The level of abstraction fits the audience and the available time.
- Things are explained on a conceptual level, not on the level of the implementation.
- The content is correct, particularly fundamental definitions.
- Terms and abbreviations are introduced or avoided (unless they are very well known), and the terminology is consistent.
- Concrete examples (e.g., an input text and the output of the model) explain the task, the method, and errors.
- The contribution is presented realistically, not overstated or oversold.
- Limitations and open questions are discussed, together with your critical assessment.

Ending

- A summary slide presents a few clear take-home messages that answer the research question.
- This summary (and your contact information) stays on screen during the discussion, instead of a slide that only says “Thank you! Questions?”.
- The talk ends on time and closes clearly, instead of trailing off (“… so yeah, that’s it”).

Slide design

- The layout is consistent and coherent and does not distract from the content.
- The font size is readable and consistent across slides.
- The colors have good contrast (no light text on a light background).
- Color coding has a meaning and is explained; it does not rely on red and green alone (color blindness).
- Bullet points are formatted consistently.
- There are no long texts to be read word by word: keywords on the slide, explanations spoken.
- The slides are numbered, so that the audience can refer to them in the discussion.
- Content in other languages is translated.
- References appear on the slides where they matter, not only at the end.
- Images contribute to the content; there are no merely decorative images.
- Images are legally allowed to be reproduced. If that is not possible, redraw them and cite the source. Avoid images with watermarks (they are also hard to read).
- There are no inside jokes; they typically do not work.

Figures, tables, equations, and code

- Flow charts and diagrams are presented in an understandable order, and the speaker walks the audience through them.
- Figures have axis labels, legends, and units. The scale of bar charts is appropriate (e.g., no truncated axis that exaggerates differences).
- Every figure is interpreted: the speaker says what the audience should see in it.
- Full tables of numbers are rarely helpful. Show only the relevant rows and columns, highlight the main result, and talk the audience through it.
- Complicated equations are rarely helpful. If they are needed, define every symbol and give the intuition first.
- Code on slides is not necessary, but sometimes helpful; if used, it is short and readable.
- No two figures show the same content.

Delivery

- The speaker faces the audience, makes eye contact, and talks to them, not to the screen.
- The speaker speaks clearly, at a calm pace, and pauses after key points.
- The speaker avoids filler words and does not read from the slides or from notes.
- The pointer is used deliberately: point once at what matters, don’t circle around.

- The speaker is polite and friendly.
- The speaker engages with the audience: asks questions and checks whether they can follow.

Time management

- The talk stays within the time limit.
- The time is distributed sensibly; the results are not rushed at the end.
- No slides are skipped during the talk; remove them beforehand (or move them to a backup section).

Discussion

- Questions are repeated or rephrased before they are answered.
- Answers are honest; it is fine to say “I don’t know” or “the paper does not say”.
- Backup slides are prepared for expected questions (e.g., details on the data or the method).

Group presentations

- The group gives one joint talk with one slide deck and a consistent style, not two separate talks.
- The talk is split into coherent parts, and all presenters have similar shares.
- The transitions are rehearsed together, and all presenters can answer questions about the whole paper.